WHAT IS CLAIMED IS:

Sub A1

1. A document type definition generating method, comprising, in a structured document provided with a tag having an element name in each document element:

a physical structure judging step of judging a physical structure of each document element;

a semantic structure judging step of judging a semantic structure of said each document element; and

a document type definition generating step of generating document type definition to define appearance state of the document element in said structured document based on judgment results of said physical structure judging step and said semantic structure judging step.

2. The document type definition generating method according to claim 1, wherein said physical structure judging step comprises judging the physical structure of the document element based on an indention or a blank line.

3. The document type definition generating method according to claim 2, wherein when the physical structure of the document element is judged based on said indention, the judging is performed by excluding the indention which represents quotation.

4. The document type definition generating method according to claim 2, wherein when the physical structure of the document element is judged based on said blank line, the judging is performed by excluding the blank line from a document in which description is made by constantly placing every predetermined number of blank lines.

5. The document type definition generating method according to claim 1, wherein said physical structure judging step comprises judging the physical structure of the document element based on a positional relation of the tags surrounding the document element.

6. The document type definition generating method according to claim 1, wherein said semantic structure judging step comprises referring to a semantic information database to judge the semantic structure of the document element based on words and phrases connection in a document and word types.

7. The document type definition generating method according to claim 1, wherein said semantic structure judging step comprises judging the semantic structure of the document element based on a meaning represented by the tags surrounding the document element.

8. The document type definition generating method according to claim 1, wherein said document type definition generating step comprises a redundancy removing step of, when the physical structure and the semantic structure of a plurality of document elements having the tags different in element name are similar, regarding the document elements as being of the same type and excluding one element name from a document type definition generating object based on the judgment results of said physical structure judging step and said semantic structure judging step.

9. The document type definition generating method according to claim 8, wherein said redundancy removing step comprises obtaining similarity degrees concerning agreement degrees of the physical structure and the semantic structure between the document elements having the tags different in element name, and regarding the document elements as being of the same type when a general similarity value calculated from the similarity degrees is equal to or more than a predetermined threshold value.

10. The document type definition generating method according to claim 1, wherein said document type definition generating step comprises a title changing step of, when the physical structure and the semantic

structure of a plurality of document elements having the tags with the same element name are different, regarding the document elements as being of different types and changing one element name based on the judgment results of said physical structure judging step and said semantic structure judging step.

11. The document type definition generating method according to claim 1, wherein said document type definition generating step comprises analyzing words and phrases present between a start tag and an end tag having the same title, obtaining information to be included between the tags, and generating the document type definition based on the information.

12. A document type definition generating apparatus comprising: in a structured document provided with a tag having an element name in each document element,

physical structure judging means for judging a physical structure of said each document element;

semantic structure judging means for judging a semantic structure of said each document element; and

document type definition generating means for generating document type definition to define appearance state of the document element in said structured document based on judgment results of said physical structure judging means and said semantic structure judging means.

13. The document type definition generating apparatus according to claim 12, wherein said physical structure judging means judges the physical structure of the document element based on an indention or a blank line.

14. The document type definition generating apparatus according to claim 13, wherein said physical structure judging means judges the physical structure of the document element based on said indention by excluding the indention which represents quotation.

15. The document type definition generating apparatus according to claim 13, wherein said physical structure judging means judges the physical structure of the document element based on said blank lines by excluding the blank lines from a document in which description is made by constantly placing every predetermined number of blank lines.

16. The document type definition generating apparatus according to claim 12, wherein said physical structure judging means judges the physical structure of the document element based on a positional relation

of the tags surrounding the document element.

17. The document type definition generating apparatus according to claim 12, wherein said semantic structure judging means refers to a semantic information database to judge the semantic structure of the document element based on words and phrases connection in a document and word types.

18. The document type definition generating apparatus according to claim 12, wherein said semantic structure judging means judges the semantic structure of the document element based on a meaning represented by the tags surrounding the document element.

19. The document type definition generating apparatus according to claim 12, wherein said document type definition generating means comprises redundancy removing means for, when the physical structure and the semantic structure of a plurality of document elements having the tags different in element name are similar, regarding the document elements as being of the same type and excluding one element name from a document type definition generating object based on the judgment results of said physical structure judging means and said semantic structure judging means.

20. The document type definition generating apparatus according to claim 19, wherein said redundancy removing means obtains similarity degrees concerning agreement degrees of the physical structure and the semantic structure between the document elements having the tags different in element name, and regards the document elements as being of the same type when a general similarity value calculated from the similarity degrees is equal to or more than a predetermined threshold value.

21. The document type definition generating apparatus according to claim 12, wherein said document type definition generating means comprises title changing means for, when the physical structure and the semantic structure of a plurality of document elements having the tags with the same element name are different, regarding the document elements as being of different types and changing one element name based on the judgment results of said physical structure judging means and said semantic structure judging means.

22. The document type definition generating apparatus according to claim 12, wherein said document type definition generating means analyzes words and phrases present between a start tag and an end tag having the same title, obtains information to be

included between the tags, and generates the document type definition based on the information.

23. A computer-readable storage medium storing a document type definition generating program for controlling a computer to perform document type definition generation, said program comprising codes for causing the computer to perform:

in a structured document provided with a tag having an element name in each document element, a physical structure judging step of judging a physical structure of each document element;

a semantic structure judging step of judging a semantic structure of said each document element; and

a document type definition generating step of generating document type definition to define appearance state of the document element in said structured document based on judgment results of said physical structure judging step and said semantic structure judging step.